REMARKS/ARGUMENTS

Claims 44-47 and 49-51 are pending in this application. The rejections to the claims are respectfully traversed.

Claim Rejections-35 U.S.C. §§101/112, First Paragraph

Claims 44-47 and 49-51 are rejected under 35 U.S.C. §101, allegedly because the claimed invention is not supported by either a specific asserted utility or a well-established utility.

Claims 44-47 and 49-51 are further rejected under 35 U.S.C. §112, first paragraph, allegedly "since the claimed invention is not supported by either a specific and substantial asserted utility or a well established utility, one skilled in the art would not know how to use the claimed invention."

For the reasons outlined below, Applicants respectfully disagree and traverse the rejection. With respect to Claims 44-47 and 49-51, Applicants submit that not only has the Patent Office not established a *prima facie* case for lack of utility and enablement, but that the PRO343 polypeptides possess a credible, specific and substantial asserted utility and are fully enabled.

Applicants have asserted utility for the instantly claimed PRO343 polypeptide based on amplification of the PRO343 gene in the "gene amplification assay" described in the instant specification in Example 92. Gene amplification is an essential mechanism for oncogene activation. It is well known that gene amplification occurs in most solid tumors, and generally is associated with poor prognosis. As described in Example 92 of the present application, the inventors isolated genomic DNA from a variety of primary cancers and cancer cell lines that are listed in Table 9 (pages 230-234 of the specification), including primary lung and colon cancers of the type and stage indicated in Table 8 (page 227). As a negative control, DNA was isolated from the cells of ten normal healthy individuals, which was pooled and used as a control (page 222, lines 34-36). Gene amplification was monitored using real-time quantitative TaqMan™ PCR. The gene amplification results are set forth in Table 9. As explained in the passage bridging pages 222 and 223, the results of TaqMan™ PCR are reported in ΔCt units. One unit corresponds to one PCR cycle or approximately a 2-fold amplification, relative to control, two units correspond to 4-fold, 3 units to 8-fold, etc. amplification. PRO343 showed

ΔCt values of approximately 1.00-3.62 in seven lung tumors and 1.15-3.49 thirteen colon tumors. This corresponds to at least **2.00- 12.3 fold amplification in lung tumors** and at least **2.22- 11.24 fold amplification in colon tumors**. Accordingly, the present specification clearly discloses strong evidence that the gene encoding the PRO343 polypeptide is significantly amplified in a significant number of lung and colon tumors.

In further support for the "significance" of the amplification, Applicants had submitted, in their Response filed March 11, 2003, a Declaration by Dr. Audrey Goddard. Applicants particularly draw the Examiner's attention to page 3 of the Goddard Declaration which clearly states that:

> It is further my considered scientific opinion that an at least **2-fold increase** in gene copy number in a tumor tissue sample relative to a normal (*i.e.*, non-tumor) sample is significant and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. Accordingly, a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is **useful as a marker for the diagnosis of cancer**, for monitoring cancer development and/or for measuring the efficacy of cancer therapy. (Emphasis added).

In addition, the Goddard Declaration clearly establishes that the TaqMan real-time PCR method described in Example 92 has gained wide recognition for its versatility, sensitivity and accuracy, and is in extensive use for the study of gene amplification. The facts disclosed in the Declaration also confirm that based upon the gene amplification results, one of ordinary skill would find it credible that PRO343 is a diagnostic marker of lung or colon cancer.

The Examiner notes that "the present claims are drawn to the polypeptide PRO343, not the polynucleotide" and therefore, there is allegedly, no specific, credible or substantial utility for the claimed polypeptides. The Examiner quotes Chen *et al.* to support this view.

Applicants submit that they had presented supportive evidence with their response mailed August 11, 2004 to show that the art generally teaches that "it is more likely than not" for amplified genes to also result in increased mRNA and protein levels. First, the articles by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.*, collectively teach that in general, gene

amplification increases mRNA expression. For instance, Orntoft *et al.* studied transcript levels of 5600 genes in malignant bladder cancers, many of which were linked to the gain or loss of chromosomal material, and found that in general (18 of 23 cases) chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Orntoft *et al.* showed a clear correlation between mRNA and protein expression levels in the proteins they studied and state that, "In general there was a highly significant correlation ($p<0.005$) between mRNA and protein alterations.... 26 well focused proteins whose genes had a known chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ($p<0.005$) with the mRNA changes detected using the arrays." (See page 42, column 2 to page 34, column 2). Accordingly, Orntoft *et al.* clearly support Applicants' position that proteins expressed by genes that are amplified in tumors are useful as cancer markers.

Similarly Hyman *et al.* compared DNA copy numbers and mRNA expression of over 12,000 genes in breast cancer tumors and cell lines, and found that there was evidence of a prominent global influence of copy number changes on gene expression levels. In Pollack *et al.*, the authors profiled DNA copy number alteration across 6,691 mapped human genes in 44 predominantly advanced primary breast tumors and 10 breast cancer cell lines, and found that on average, a 2-fold change in DNA copy number was associated with a corresponding 1.5-fold change in mRNA levels. In summary, the evidence supports the Appellants' position that gene amplification is more likely than not predictive of increased mRNA and polypeptide levels.

Second, the Declaration of Dr. Paul Polakis, principal investigator of the Tumor Antigen Project of Genentech, Inc., the assignee of the present application, shows that, in general, there is a correlation between mRNA levels and polypeptide levels.

Applicants further submit that, contrary to the Examiner's assertion, the cited Chen *et al.* reference does not conclusively establish a *prima facie* case for lack of utility for the PRO343 polypeptide. For instance, Applicants note that the proteins selected for their study in Chen *et al.* were identified by staining of 2D gels. As is well known, there are problems with selecting proteins detectable by 2D gels: "It is apparent that without prior enrichment only a relatively small and highly selected population of *long-lived, highly expressed proteins* is observed. There are many more proteins in a given cell which are not visualized by such methods. Frequently it is

the low abundance proteins that execute key regulatory functions" (page 1870, col. 1). Thus, Chen *et al.*, by selecting proteins visualized by 2D gels, are likely to have excluded in their analysis many key regulatory proteins which could be candidate cancer markers.

Secondly, the manner in which the Chen data was averaged and analyzed is a vastly different manner from that of the instant specification. For example, Chen *et al.* studied expression levels across a set of samples which included a large number of tumor samples (76) and a much smaller group of normal samples (9). The authors determined the global relationship between mRNA and corresponding protein expression using the average expression values for all 85 lung tissue samples. The authors chose an arbitrary threshold of 0.115 for the correlation to be considered significant. This resulted in negative normalized protein values in some cases and the authors concluded that it is not possible *to predict* overall protein expression based on **average mRNA abundance**. Once again, Applicants remind the Examiner that the utility standard does not require accurate prediction of protein values; only that in a majority of the proteins studied, it is more likely than not that protein levels increased when mRNA levels increased. A review of the correlation coefficient data presented in the Chen *et al.* paper indicates that, in fact, Chen teaches that 'it is more likely than not' that increased mRNA expression correlates well with increased protein expression. For instance, a review of Table 1, which lists 66 genes [the paper incorrectly states there are 69 genes listed] for which only one protein isoform is expressed, shows that 40 genes out of 66 had a positive correlation between mRNA expression and protein expression. This clearly meets the test of "more likely than not". Similarly, in Table II , 30 genes with multiple isoforms [again the paper incorrectly states there are 29] were presented. In this case, for 22 genes out of 30, at least one isoform showed a positive correlation between mRNA expression and protein expression. Furthermore, 12 genes out of 29 showed a strong positive correlation [as determined by the authors] for at least one isoform. No genes showed a significant negative correlation. It is not surprising that not all isoforms are positively correlated with mRNA expression. Thus, Table II also provides that it is more likely than not that protein levels will correlate with mRNA expression levels.

The same authors in Chen *et al.*, published a later paper, Beer *et al.*, Nature Medicine 8(8) 816-824 (2002) (copy enclosed) which described gene expression of genes in adenocarcinomas

and compared that to protein expression. In this paper they report that "these results suggest that the oligonucleotide microarrays provided reliable measures of gene expression." (pg 317). The authors also state "these studies indicate that many of the genes identified using gene expression profiles are likely relevant to lung adenocarcinoma." Clearly the authors of the Chen paper agree that microarrays provide a reliable measure of the expression levels of the gene and can be used to identify genes whose overexpression is associated with tumors.

As was discussed in the Utility standard submitted in previous responses, the law does not require the existence of a "strong" or "linear" correlation between mRNA and protein levels. Nor does the law require that protein levels be "accurately" predicted. Accordingly, the data by Chen *et al.* confirm that there is a general trend between protein expression and transcript levels, which meets the "more likely than not standard" and shows that a positive correlation exists between mRNA and protein. Therefore, Applicants submit that the Examiner's Utility rejection is based on a misrepresentation of the scientific data presented in Chen *et al.* and by applying an improper, heightened legal standard in this case. In fact, contrary to what the Examiner contends, the art indicates that, if a gene is amplified in cancer, it is more likely than not that the mRNA and the encoded protein will also be expressed at an elevated level. As noted even in Chen *et al.* most genes showed a correlation between increased mRNA and translated protein.

The Examiner also notes that "the claimed sequences merely revealed similarity to proteases in genenral" (page 3, last paragraph of Office action). Applicants respectfully assert that utility for the instant PRO343 is based on the results in the gene amplification assay, not on structure prediction, and hence such a rejection is moot.

Taken together, although there are some examples in the scientific art that do not fit within the central dogma of molecular biology that there is a correlation between DNA, mRNA, and polypeptide levels, these instances are exceptions rather than the rule. In the majority of amplified genes, as exemplified by Ornoft *et al.*, Hyman *et al.*, Pollack *et al.*, the Polakis Declaration and the widespread use of array chips, the teachings in the art overwhelmingly show that gene amplification influences gene expression at the mRNA and protein levels. Therefore, one of skill in the art would reasonably expect in this instance, based on the amplification data for the PRO343 gene, that the PRO343 polypeptide is concomitantly overexpressed.

Thus, Applicants have demonstrated utility for the PRO343 polypeptide based on the gene amplification assay and thus, Applicants request that the Examiner reconsider the utility for the present application based on the present arguments. Furthermore, since the specification has provided detailed protocols for the gene amplification assay, for example, in Example 92, one of ordinary skill in the art could identify that the claimed polypeptides could be made and used in the diagnosis of lung or colon tumors, without any undue experimentation.

Hence Applicants respectfully request reconsideration and reversal of the utility/ enablement rejection of the pending claims under 35 U.S.C. §§101/112, first paragraph.

The present application is believed to be in *prima facie* condition for allowance, and an early action to that effect is respectfully solicited.

Please charge any additional fees, including any fees for additional extension of time, or credit overpayment to Deposit Account No. **08-1641** (referencing Attorney's Docket No. **39780-1618 P2C48**). Please direct any calls in connection with this application to the undersigned at the number provided below.

Respectfully submitted,

Date: November 9, 2005

By:_____
Panpan Gao (Reg. No. 43,626)

HELLER EHRMAN, LLP
275 Middlefield Road
Menlo Park, California 94025
Telephone: (650) 324-7000
Facsimile: (650) 324-0638

SV 2166627 v1
11/9/05 11:15 AM (39780.1618)

Gene-expression profiles predict survival of patients with lung adenocarcinoma

DAVID G. BEER[1], SHARON L.R. KARDIA[2], CHIANG-CHING HUANG[3], THOMAS J. GIORDANO[4], ALBERT M. LEVIN[2], DAVID E. MISEK[5], LIN LIN[1], GUOAN CHEN[1], TAREK G. GHARIB[1], DAFYDD G. THOMAS[4], MICHELLE L. LIZYNESS[4], RORK KUICK[5], SATORU HAYASAKA[3], JEREMY M.G. TAYLOR[3], MARK D. IANNETTONI[1], MARK B. ORRINGER[1] & SAMIR HANASH[5]

Departments of [1]Surgery, [2]Epidemiology, [3]Biostatistics, [4]Pathology and [5]Pediatrics, University of Michigan, Ann Arbor, Michigan, USA
Correspondence should be addressed to D.G.B.; email: dgbeer@umich.edu.

Published online: 15 July 2002, doi:10.1038/nm733

Histopathology is insufficient to predict disease progression and clinical outcome in lung adenocarcinoma. Here we show that gene-expression profiles based on microarray analysis can be used to predict patient survival in early-stage lung adenocarcinomas. Genes most related to survival were identified with univariate Cox analysis. Using either two equivalent but independent training and testing sets, or 'leave-one-out' cross-validation analysis with all tumors, a risk index based on the top 50 genes identified low-risk and high-risk stage I lung adenocarcinomas, which differed significantly with respect to survival. This risk index was then validated using an independent sample of lung adenocarcinomas that predicted high- and low-risk groups. This index included genes not previously associated with survival. The identification of a set of genes that predict survival in early-stage lung adenocarcinoma allows delineation of a high-risk group that may benefit from adjuvant therapy.

Lung cancer remains the leading cause of cancer death in industrialized countries. Most patients with non-small cell lung cancer (NSCLC) present with advanced disease, and despite recent advances in multi-modality therapy, the overall 10-year survival rate remains a dismal 8–10%[1]. However, a significant minority of patients (~25–30%) with NSCLC have stage I disease and receive surgical intervention alone. Although 35–50% of patients with stage I disease will relapse within 5 years[2–4], it is not currently possible to identify specific high-risk patients.

Adenocarcinoma is currently the predominant histological subtype of NSCLC (refs. 1,5,6). Although morphological assessment of lung carcinomas can roughly stratify patients, there is a need to identify patients at high risk for recurrent or metastatic disease. Preoperative variables that affect survival of patients with NSCLC have been identified[7–10]. Tumor size, vascular invasion, poor differentiation, high tumor-proliferative index and several genetic alterations, including K-*ras* (refs. 11,12) and p53 (refs. 10,13) mutations, have prognostic significance. Multiple independently assessed genes or gene products have also been investigated to better predict patient prognosis in lung cancer[14–18]. Technologies that simultaneously analyze the expression of thousands of genes[19] can be used to correlate gene-expression patterns with numerous clinical parameters—including patient outcome—to better predict tumor behavior in individual patients[20]. Analyses of lung cancers using array technologies have identified subgroups of tumors that differ according to tumor type and histological subclasses and, to a lesser extent, survival among adenocarcinoma patients[21,22]. Here we correlated gene-expression profiles with clinical outcome in a cohort of patients with lung adenocarcinoma and identified specific genes that predict survival among patients with stage I disease. For further validation, we also show that the risk index predicted survival in an independent cohort of stage I lung adenocarcinomas.

Hierarchical profile clustering yields three tumor subsets

Using oligonucleotide arrays, we generated gene-expression profiles for 86 primary lung adenocarcinomas, including 67 stage I and 19 stage III tumors, as well as 10 non-neoplastic lung samples. Selected sample replicates showed high correlation among coefficients and reliable reproducibility. We determined transcript abundance using a custom algorithm and the data set was trimmed of genes expressed at extremely low levels, that is, genes were excluded if the measure of their 75th percentile value was less than 100. Although potentially resulting in the loss of some information, trimming in this manner decreased the possibility that the clustering algorithm would be strongly influenced by genes with little or no expression in these samples. Hierarchical clustering with the resulting 4,966 genes yielded 3 clusters of tumors (Fig. 1). All 10 non-neoplastic samples clustered tightly together within Cluster 1 (data not shown). We examined the relationships between cluster and patient and tumor characteristics (Fig. 1 and Supplementary Figure A online). There were associations between cluster and stage ($P = 0.030$) and between cluster and differentiation ($P = 0.01$). Cluster 1 contained the greatest percentage (42.8%) of well differentiated tumors, followed by Cluster 2 (27%) and Cluster 3 (4.7%). Cluster 3 contained the highest percentage of both poorly differentiated (47.6%) and stage III tumors (42.8%), yet contained 3 (14.3%) moderately differentiated and 1 (5%) well differentiated stage I tumor. Notably, 11 stage I tumors were present in Cluster 3, sug-

gesting a common gene-expression profile for this subset of stage I and stage III tumors.

For patients with stage I and stage III tumors, the average ages were 68.1 and 64.5 years and the percentage of smokers was 88.9% and 89.5%, respectively. Marginally significant associations between cluster and smoking history were observed ($P = 0.06$). A significant relationship between histopathological classification and cluster was only discernable for bronchioloalveolar adenocarcinomas (BAs), which were only present in Clusters 1 and 2 ($P = 0.0055$) and comprised 35.7% and 12.3% of tumors for Clusters 1 and 2, respectively.

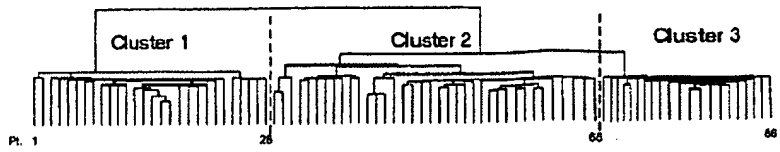


Fig. 1 Unsupervised classification analysis of lung adenocarcinomas. 3 classes of tumors identified by agglomerative hierarchical clustering of gene-expression profiles using the 4,966 expressed genes. Patient and histopathological information for each lung adenocarcinoma case by cluster designation and methods for K-*ras* 12/13th-codon mutational status and nuclear p53 protein accumulation are provided (Supplementary Figure A online). TN classification denotes information regarding patient tumor size and nodal involvement. Associations between cluster membership and patient or histopathological variables are indicated at significance level ($P \leq 0.05$).

We examined the heterogeneity in gene-expression profiles based on the trimmed data set among normal lung samples and stage I and stage III adenocarcinomas by calculating correlation coefficients between all pairs of samples. In contrast to normal lung samples that displayed highly similar gene-expression profiles (median correlation, 0.9), both stage I and III lung tumors demonstrated much greater heterogeneity in their expression profiles with lower correlation coefficients (median values, 0.82 and 0.79, respectively).

Northern-blot and immunohistochemistry analyses

Of the 4,966 genes examined, 967 differed significantly between stage I and III adenocarcinomas, a number in excess of that expected by chance alone (248 at alpha level (α) = 0.05). Three genes were arbitrarily selected to verify the microarray expression data. The mRNA from 20 of the normal lung and tumor samples was examined by northern-blot hybridization with probes for insulin-like growth factor–binding protein 3 (*IGFBP3*), cystatin C and lactate dehydrogenase A (*LDH-A*) (Fig. 2*a*). Two gene probes not represented on the microarrays were used as controls, including histone H4, a potential index of overall cell proliferation, and 28S ribosomal RNA, a control for sample loading and transfer. The relative amounts of *IGFBP3*, cystatin C and *LDH-A* mRNA strongly correlated with microarray-based measurements (Fig. 2*b*). In both assays, *IGFBP3* and *LDH-A* mRNA levels increased from stage I to stage III adenocarcinomas and were higher than those in normal lung. Cystatin C mRNA levels were more variable but relatively greater in normal lung than tumors. These results suggest that the oligonucleotide microarrays provided reliable measures of gene expression. The tumors showed slightly greater histone H4 expression than the normal lung, likely reflecting increased proliferation of tumor cells.

Immunohistochemistry was performed for IGFBP3, cystatin C and HSP-70 to determine whether mRNA overexpression was reflected by an increase of their corresponding proteins in tumors.

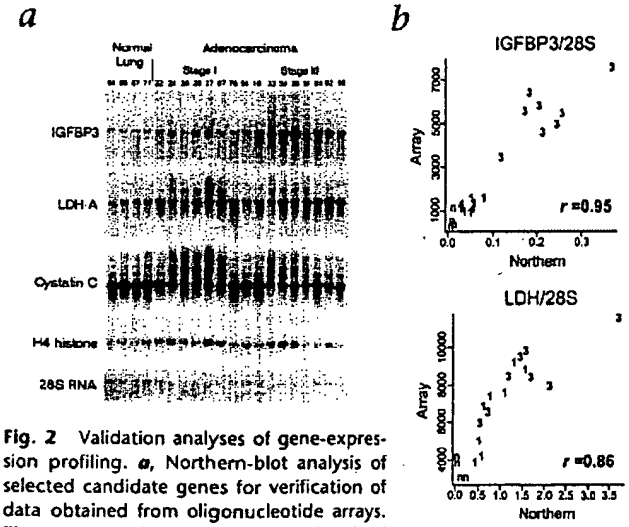
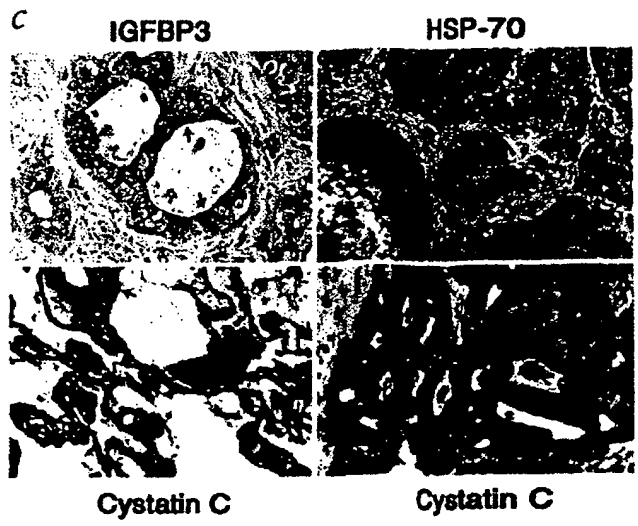


Fig. 2 Validation analyses of gene-expression profiling. ***a***, Northern-blot analysis of selected candidate genes for verification of data obtained from oligonucleotide arrays. The same sample RNA for the 4 uninvolved lung, 8 stage I and 8 stage III tumors was used for the northern-blot and oligonucleotide array analyses. ***b***, Correlation analysis of quantitative data obtained from oligonucleotide arrays and northern blots measured by integrated phosphorimager-based signals for the *IGFBP3* and *LDH-A* genes. The ratio of *IGFBP3*, cystatin C and *LDH-A* mRNA to *28S* rRNA was determined. The relative values for each gene from each sample are shown. n, non-neoplastic normal lung; 1, stage I tumors; 3, stage III tumors. ***c***, Immunohistochemical analysis of IGFBP-3, HSP-70 and cystatin C in lung and lung adenocarcinomas. Cytoplasmic IGFBP-3 immunoreactivity in a neoplastic gland (tumor L22) with prominent apical staining (blue reactant staining, arrow, upper left). Diffuse cytoplasmic HSP-70 immunoreactivity (tumor L27), yet stromal elements show no reactivity (upper right). Normal lung parenchyma (lower left) shows cytoplasmic cystatin C immunoreactivity in alveolar pneumocytes (arrow) and intra-alveolar macrophages but tumor (L90) shows diffuse cytoplasmic cystatin C immunoreactivity with prominent apical staining (lower right). Magnification, ×200



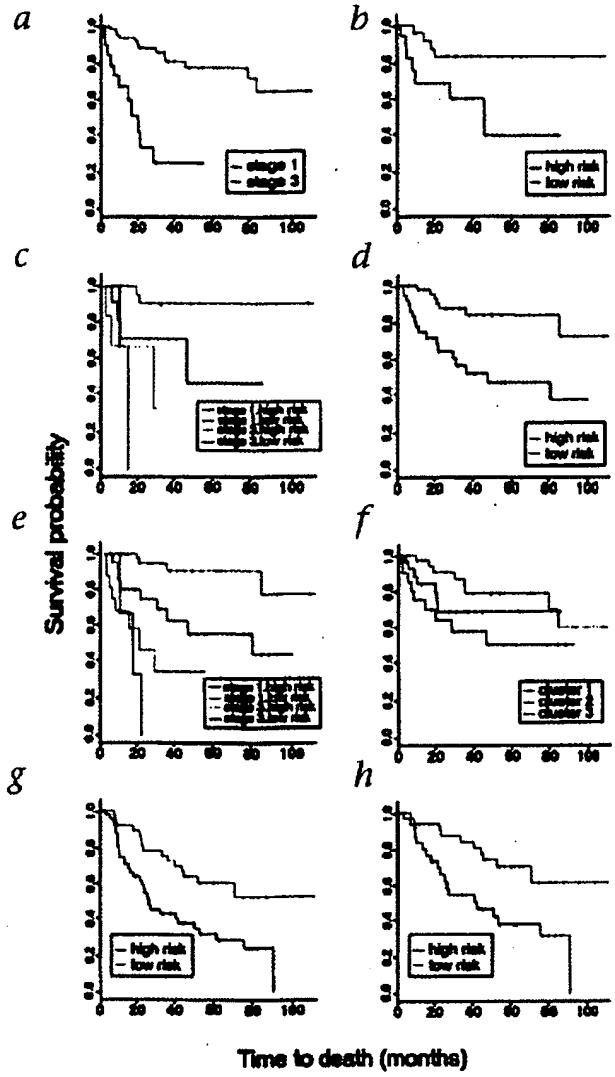


Fig. 3 Gene-expression profiles and patient survival. *a*, Relationship between tumor stage and patient survival (stage 1 and stage 3 differ significantly, $P < 0.0001$). *b*, Relationship between the survival in the 43 test samples and their risk assignments based on the 50-gene risk index estimated in the 43 training samples. The high- and low-risk groups differ significantly ($P = 0.024$). *c*, Relationship between patient survival and the risk assignments in test samples (in *b*) conditional for tumor stage. The high- and low-risk stage I groups differ significantly ($P = 0.028$), whereas stage III low- and high-risk groups did not ($P = 0.634$). *d*, Relationship between survival in the test cases and their risk assignments based on the 86 'leave-one-out' cross-validation of the 50-gene risk index. The high- and low-risk groups differ significantly ($P = 0.0006$). *e*, Relationship between test case's risk assignment and survival (in *d*) conditional on tumor stage. The high- and low-risk stage I lung adenocarcinoma groups differ significantly from each other ($P = 0.003$), whereas low- and high-risk stage III tumors do not. *f*, Relationship between tumor class identified by hierarchical clustering and patient survival. Survival for patients in Cluster 3 differed relative to the tumors in Cluster 2 ($P = 0.037$) and approached significance for Cluster 1 and 2 combined ($P = 0.06$). *g*, Analysis of the Michigan-based risk index using top cross-validated survival genes identify a low- and high-risk group in an independent cohort of 84 Massachusetts-based lung adenocarcinomas that are significantly different ($P = 0.003$). *h*, Among the 62 stage I lung adenocarcinomas in the Massachusetts sample, the high- and low-risk groups differed significantly ($P = 0.006$).

Immunoreactivity for both *IGFBP-3* and *HSP-70* (Fig. 2*c*) was detected in the cytoplasm of the adenocarcinomas, with little detectable reactivity in the stromal or inflammatory cells. Cystatin C was detected in alveolar pneumocytes and intra-alveolar macrophages in non-neoplastic lung parenchyma and also consistently in the cytoplasm of neoplastic cells.

Gene-expression profiles predict survival

As expected, Kaplan–Meier survival curves (Fig. 3*a*) and log-rank tests indicated poorer survival among stage III compared with stage I adenocarcinomas ($P = <0.0001$). Two statistical approaches were used to determine whether gene-expression profiles could predict survival using the data set of 4,966 genes. In one approach, equal numbers of randomly assigned stage I and stage III tumors constituted training ($n = 43$) and testing ($n = 43$) sets. In the training set, the top 10, 20, 50 or 75 genes were used to create risk indices that were evaluated for their association with survival using the 50th, 60th or 70th percentile cutoff points to categorize patients into high or low groups. The results were similar across cutoff points but the 50-gene risk index had the best overall association with survival in the training set. After conservatively choosing the 60th percentile cutoff point from the training set, we then applied this risk index and cutoff point to the testing set. The risk index of the top 50 genes correctly identified low- and high-risk individuals within the independent testing set ($P = 0.024$) (Fig. 3*b* and Supplementary Methods online). Notably, 11 stage I tumors were included in the high-risk subgroup. When this risk assignment was then conditionally examined for stage progression (Fig. 3*c*), low- and high-risk groups among stage I tumors were found to differ ($P = 0.028$) in their survival.

Identification of a robust set of survival genes

Although predictive of patient survival, a single training-testing set may not provide the most robust set of genes due to random sampling issues. Therefore, a 'leave-one-out' cross-validation approach was used to identify genes associated with survival from all 86-tumor samples. We first developed a 50-gene risk index in each training set, and then applied the risk index to the test case held out from the full set of tumors and assigned the held out tumor to the high- or low-risk groups (Fig. 3*d*). The high and low-risk subgroups determined in the test cases differed significantly in their overall survival ($P = 0.0006$). Among the larger group of stage I lung adenocarcinomas, the low-risk ($n = 46$) and high-risk ($n = 21$) groups had markedly different survival ($P = 0.003$) (Fig. 3*e*). Table 1 lists selected examples of the cumulative top 100 genes derived from this cross-validation procedure (complete list in Supplementary Table A online).

It was also noted that many of the stage I patients in the high-risk subgroup (Fig. 3*e*) were present in Cluster 3 (Fig. 1). Kaplan–Meier analysis (Fig. 3*f*) demonstrated a significantly worse survival ($P = 0.037$) for patients in Cluster 3 relative to patients in Cluster 2 and approaching significance for Cluster 1 and 2 combined ($P = 0.06$). This further indicates the important relationship between gene-expression profiles and patient survival, independent of disease stage.

Consistent with previous analyses of lung adenocarcinomas[23], 40% of stage I and 57.8% of stage III tumors had 12th or 13th codon K-*ras* gene mutations. Those patients with tumors containing K-*ras* mutations showed a trend of poorer survival, but

Table 1 Selected examples of the top 100 genes from cross-validation

Gene name	*P* (normal versus tumor t-test)	% Change in tumor	*P* (stage I versus stage III t-test)	% Change in stage III	Coefficient β	Unigene comment
						Apoptosis-related
CASP4	0.56	-6%	0.02	57%	0.0022	Caspase 4, apoptosis-related cysteine protease
P63	9.73E-04	37%	0.03	43%	0.0010	Transmembrane protein (63 kD), endoplasmic reticulum/ Golgi intermediate compartment
						Cell adhesion and structure
KRT7	8.02E-08	126%	0.11	55%	0.0003	Keratin 7
LAMB1	0.14	-20%	0.01	60%	0.0027	Laminin, β1
						Cell cycle and growth regulators
BMP2	0.54	-21%	0.27	47%	0.0044	Bone morphogenetic protein 2
CDC6	1.31E-05	1070%	0.05	148%	0.0124	CDC6 (cell division cycle 6, *Saccharomyces cerevisiae* homolog)
S100P	2.10E-08	1572%	0.19	77%	0.0001	S100 calcium-binding protein P
SERPINE1	2.89E-03	72%	0.25	30%	0.0008	Serine (or cysteine) proteinase inhibitor, clade E (nexin)
STX1A	8.65E-08	54%	0.07	26%	0.0031	Syntaxin 1A (brain)
						Cell signaling
ADM	0.05	39%	0.04	117%	0.0016	adrenomedullin
AKAP 12	8.53E-03	-47%	0.05	214%	0.0010	A kinase (PRKA) anchor protein (gravin) 12
ARHE	0.06	-39%	0.05	87%	0.0092	ras homolog gene family, member E
GRB7	2.02E-03	38%	0.63	15%	0.0030	Growth factor receptor-bound protein 7
VEGF	6.50E-08	174%	0.02	85%	0.0013	Vascular endothelial growth factor
WNT10B	0.05	31%	0.48	20%	0.0022	Wingless-type MMTV integration site family, member 10B
						Chaperones
HSPA8	0.36	8%	9.01E-04	51%	0.0008	Heat-shock 70 kD protein 8
						Receptors
ERBB2	0.04	92%	0.37	120%	0.0013	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2
FXYD3	0.10	111%	0.31	73%	0.0046	FXYD domain-containing ion transport regulator 3
SLC20A1	1.34E-03	58%	0.02	66%	0.0021	Solute carrier family 20 (phosphate transporter), member 1
						Enzymes, cellular metabolism
CSTB	1.57E-04	50%	0.15	34%	0.0001	Cystatin B (stefin B)
CTSL	0.48	-10%	0.03	67%	0.0007	Cathepsin L
CYP24	3.16E-06	N/A	0.97	2%	0.0008	Cytochrome P450, subfamily XXIV (vitamin D 24-hydroxylase)
FUT3	1.07E-07	114%	0.97	-1%	0.0033	Fucosyltransferase 3 (galactoside 3(4)-L-fucosyltransferase, Lewis blood group included)
MLN64	0.20	32%	0.42	80%	0.0007	Steroidogenic acute regulatory protein related
PDE7A	0.12	33%	0.01	-35%	-0.0187	Phosphodiesterase 7A
PLGL	0.04	-68%	0.35	-170%	-0.0011	Plasminogen-like
SLC1A6	0.07	-32%	0.12	86%	0.0069	Solute carrier family 1 (high-affinity aspartate/ glutamate transporter), member 6
						Transcription and translation
COPEB	0.10	-33%	0.26	25%	0.0016	Core promoter element binding protein
CRK	0.10	32%	0.03	48%	0.0098	v-crk avian sarcoma virus CT10 oncogene homolog
RELA	0.26	-7%	0.01	20%	0.0034	v-rel avian reticuloendotheliosis viral oncogene homolog A
						Unknown function
KIAA0005	2.21E-04	40%	0.02	45%	0.0010	KIAA0005 gene product
MGB1	0.27	125%	0.33	459%	0.0018	Mammaglobin 1

Bolded genes were also significant for survival in 43 tumor training set (Fig. 3*b*).

Table 1 Selected examples of the cumulative top 100 genes identified using training-testing, cross-validation of all 86 lung tumor samples. The percent change, as well as the direction, for the average values of the 10 non-neoplastic lung to all tumors, and for the 67 stage I to the 19 stage III tumors are shown. A positive coefficient β value is indicative of a relationship of gene expression to a poorer patient outcome. The genes are listed in potential functional categories. Genes that were also present in the top 50 survival genes using the 43-tumor training set (Fig. 3*b*) are indicated in bold type. Complete listing of the gene probe sets and annotated gene and unigene identifiers can be found in the Supplementary Methods.

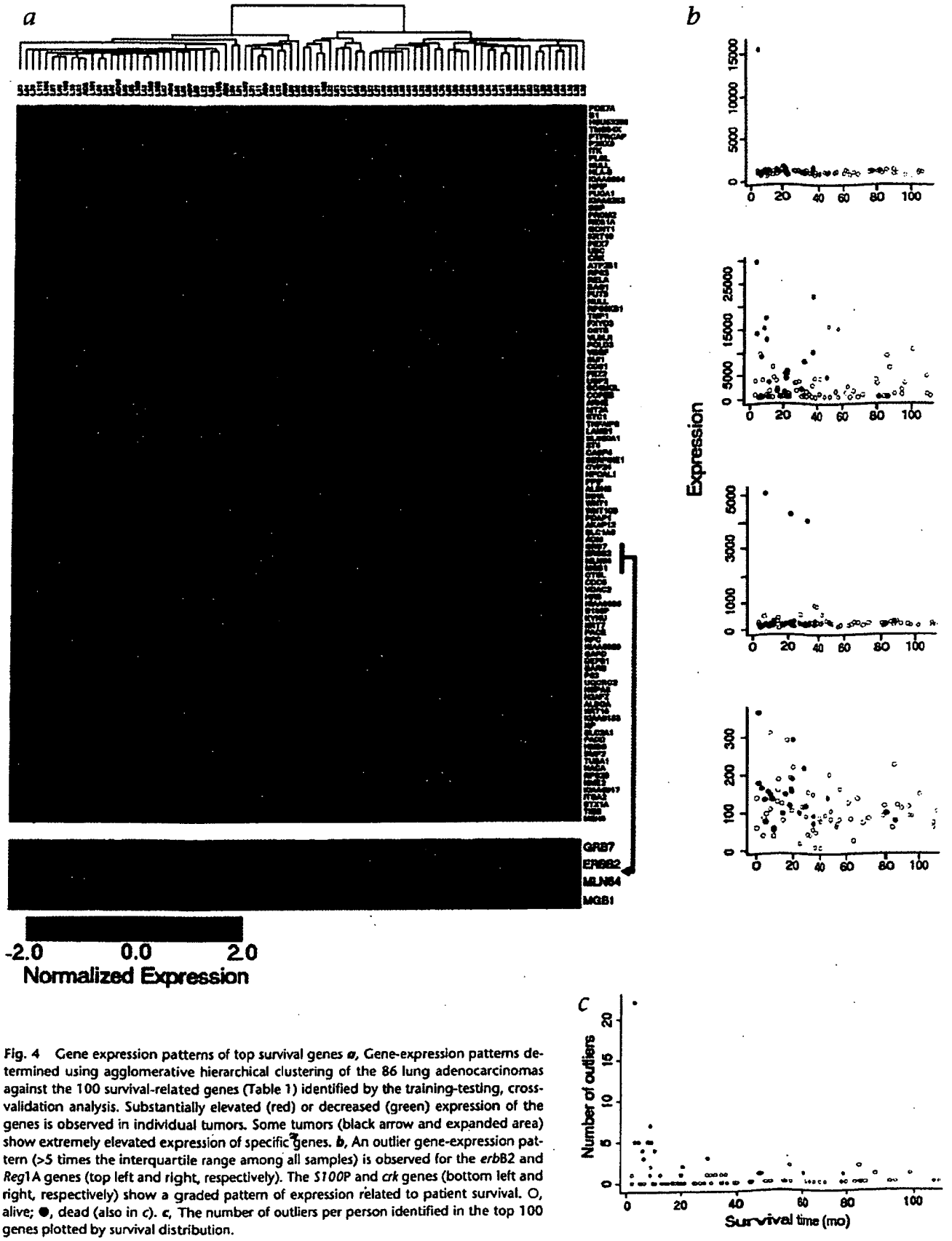


Fig. 4 Gene expression patterns of top survival genes *a*, Gene-expression patterns determined using agglomerative hierarchical clustering of the 86 lung adenocarcinomas against the 100 survival-related genes (Table 1) identified by the training-testing, cross-validation analysis. Substantially elevated (red) or decreased (green) expression of the genes is observed in individual tumors. Some tumors (black arrow and expanded area) show extremely elevated expression of specific genes. *b*, An outlier gene-expression pattern (>5 times the interquartile range among all samples) is observed for the *erbB2* and *Reg1A* genes (top left and right, respectively). The *S100P* and *crk* genes (bottom left and right, respectively) show a graded pattern of expression related to patient survival. O, alive; ●, dead (also in *c*). *c*, The number of outliers per person identified in the top 100 genes plotted by survival distribution.

Fig. 4 Gene expression patterns of top survival genes *a*, Gene-expression patterns determined using agglomerative hierarchical clustering of the 86 lung adenocarcinomas against the 100 survival-related genes (Table 1) identified by the training-testing, cross-validation analysis. Substantially elevated (red) or decreased (green) expression of the genes is observed in individual tumors. Some tumors (black arrow and expanded area) show extremely elevated expression of specific genes. *b*, An outlier gene-expression pattern (>5 times the interquartile range among all samples) is observed for the *erbB2* and *Reg1A* genes (top left and right, respectively). The *S100P* and *crk* genes (bottom left and right, respectively) show a graded pattern of expression related to patient survival. ○, alive; ●, dead (also in *c*). *c*, The number of outliers per person identified in the top 100 genes plotted by survival distribution.

this difference did not reach statistical significance among all patients ($P = 0.25$), between patients within tumor clusters ($P = 0.41$) or when analyzed separately among stage I ($P = 0.22$) and stage III ($P = 0.53$) patients. Nuclear accumulation of p53 was detected in 17.9% stage I and in 22.2% stage III tumors. No significant relationship was observed for p53 staining and patient survival, cluster or tumor stage.

Confirmation using an independent set of adenocarcinomas

The robustness of our 50-gene risk index in predicting survival in lung adenocarcinomas was tested using oligonucleotide gene-expression data obtained from a completely independent (Massachusetts-based) sample of 84 lung adenocarcinomas (62 stage I, 14 stage II and 8 stage III; ref. 21, and dataset A at www.genome.wi.mit.edu/MPR/lung). To ensure equivalent power for testing and comparability of samples, the criteria for including tumors in the analysis were 40% or greater tumor cellularity, no mixed histology (that is, adenosquamous) and patient survival information. To obtain comparative gene-expression measures between the two data sets, gene sequences present on the U95A and HuGeneFL array were examined, and expression data for our top 50 cross-validation genes for all 84 Massachusetts samples were obtained and processed[24] (see also Supplementary Methods online). When we examined the risk assignment of these 84 samples, employing the identical cutoff point used for the 86 Michigan-based lung samples, we observed low- and high-risk groups (Fig. 3*g*; $P = 0.003$). Notably, among the 62 stage I tumors, high- and low-risk groups were observed that differed significantly ($P = 0.006$) in their survival (Fig. 3*h*).

Survival genes had graded and outlier expression patterns

A statistical and graphical analysis of the 100 survival-related genes (Table 1) clustered against all 86 tumors revealed individual tumors with substantially elevated expression in both a limited and larger number of genes (Fig. 4*a*). Among these genes, we observed two distinct patterns of expression related to patient survival. One pattern, designated 'outlier', included genes showing substantially elevated expression (greater than five times the interquartile range among all samples), whereas the other pattern, designated 'graded', was characterized by continuously distributed expression with patient survival (Fig. 4*b*). The *erbB2* and *Reg1A* genes are examples of outlier expression patterns and *S100P* and *crk* genes of graded patterns. The number of outliers per person in the top 100 genes was identified and plotted according to survival times and events (Fig. 4*c*). Both stage I and stage III lung adenocarcinomas showed outlier gene patterns and 10 tumors contained 3 or more outlier genes.

Because gene amplification may result in increased gene expression, the nine genes with outlier expression patterns (*erbB2*, *SLC1A6*, *Wnt 1*, *MGB1*, *Reg1A*, *AKAP12*, *PACE*, *CYP24*, *KYNU*) and one gene with a graded expression pattern (*KRT18*) were examined using quantitative genomic PCR to evaluate genomic copy number (Fig. 5*a*). Gene amplification of *erbB2* (17q12) was detected in tumor L94, which had the highest *erbB2* mRNA expression (Fig. 4*a*). Gene amplification was not detected for any of the other seven tested genes in tumor L94, as well as in other tumors. The two genes most frequently demonstrating the outlier pattern in these lung adenocarcinomas were *KYNU* and *CYP24*, and were present in 10 and 9 tumors, respectively. *CYP24* has been described as a gene amplified and overexpressed in breast cancer[25], and these results indicate elevated expression in lung adenocarcinoma.

To determine whether the graded or outlier gene-expression patterns also occur at the protein-expression level, 10 of the 100

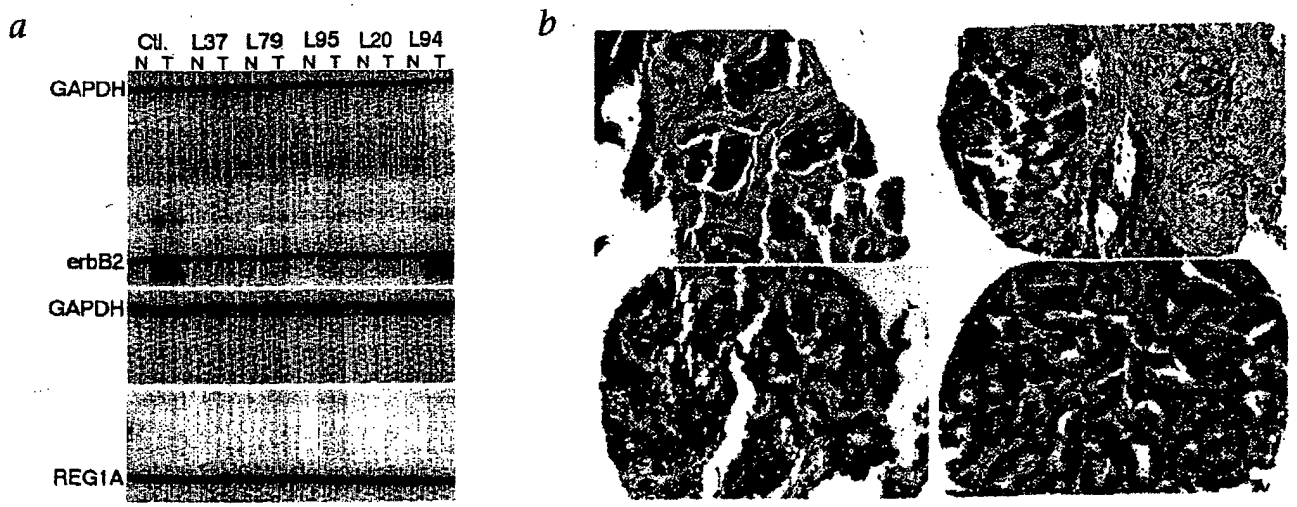


Fig. 5 Gene amplification and protein expression of survival-related genes. *a*, Analysis of potential gene amplification for 9 genes showing outlier expression patterns in the lung tumors (*erbB2*, *SLC1A6*, *Wnt 1*, *MGB1*, *Reg1A*, *AKAP12*, *PACE*, *CYP24* and *KYNU*) and examined using quantitative genomic PCR. A gene showing graded expression pattern (*KRT18*), and one gene (*PACE4*) with a similar chromosome location as *PACE*, were used as controls. Only *erbB2* and *Reg1A* are shown. An esophageal adenocarcinoma with known high-level genomic amplification of *erbB2* was used as a positive control and normal esophagus DNA was used as a negative control (Ctl). PCR fragments sizes were 343 bp for *GAPDH*, 166 bp for *erbB2* and 126 bp for *Reg1A*. DNA is from normal lung (N) and tumor(T) from each patient (for example L37). *b*, Immunohistochemical analysis of survival related genes with lung adenocarcinoma microarrays using the tumors from this study. The transmembrane erbB2 protein (top left) expression is substantially increased in tumor L94 containing the amplified *erbB2* gene (Fig. 4*a* and *b*). Expression of VEGF (top right) and S100P (bottom left) was located within the neoplastic cells and the pattern of immunoreactivity was consistent with the graded expression pattern demonstrated by their mRNA profiles. Expression of the oncogene *crk* (bottom right) was abundantly expressed in neoplastic lung cells. Magnification, ×400 (erbB2); ×200 (VEGF, S100P and crk).

top survival genes (Table 1) for which specific antibodies were available were chosen for immunohistochemical analysis using lung-tumor arrays from this study (Fig. 5*b*). Expression of membrane erbB2 protein was substantially increased in the *erb*B2-amplified tumor L94 and very low levels of expression were present in other tumors, consistent with mRNA-expression measurements (Fig. 4*a* and *b*). CDC6 protein expression was also substantially higher in tumor L94, consistent with mRNA levels (data not shown). Expression of vascular endothelial growth factor (VEGF) and S100P (Fig. 5*b*), as well as cytokeratin 18 (KRT18), cytokeratin 7 (KRT7) and fas-associated death domain (FADD) protein (data not shown), was located within the lung tumor cells and consistent with the graded expression pattern of the mRNA profiles. The oncogene *crk* showed both graded mRNA as well as a graded protein-expression pattern with survival, and was abundantly expressed in the tumor cells (Fig. 5*b*). These results indicate that many survival-associated genes are expressed at the protein level and demonstrate similar mRNA and protein-expression patterns.

Discussion

We used several approaches for the analysis of gene-expression data related to clinicopathological variables and patient survival. One approach, hierarchical clustering, was used to examine similarities among lung adenocarcinomas in their patterns of gene expression. Previous studies of lung tumors[21,22] have also used this method to describe subclasses of lung tumors. Here, we found three clusters that showed significant differences with respect to tumor stage and tumor differentiation. This suggests, as expected, that tumors with similar histological features of differentiation demonstrate similarities in gene expression. This feature also partly underlies the observed statistical association of tumor stage and cluster, as many of the higher-stage tumors, often poorly differentiated and previously associated with a reduced survival[9,10], were located in Cluster 3. Although this cluster contained the highest percentage of stage III tumors, it also contained a nearly equal mixture of stage I and stage III tumors and not all tumors were poorly differentiated. This indicates that a subset of stage I lung adenocarcinomas share gene-expression profiles with higher-stage tumors. Notably, 10 of the 11 stage I tumors found in Cluster 3 were the high-risk stage I tumors identified using the risk index in the 'leave-one-out' cross-validation.

In contrast to previous analyses of lung adenocarcinomas[21,22], we validated the expression data from the arrays. The strong correlation of northern-blot analysis and oligonucleotide-array data for gene expression in the same samples (Fig. 2*b*) indicates that these studies provide robust gene-expression estimates. Immunohistochemistry using the same tumor samples in tissue arrays demonstrates protein expression within the lung tumor cells. Together, these studies indicate that many of the genes identified using gene-expression profiles are likely relevant to lung adenocarcinoma. For example, *IGFBP3* gene expression is increased in lung adenocarcinomas (Fig. 2*c*). IGFBP3 protein modulates the autocrine or paracrine effects of insulin-like growth factors, elevated IGFBP3 expression is observed in colon cancer[26], and increased serum IGFBP3 is associated with progression in breast cancer[27]. Heat-shock protein 70 (HSP-70) is increased in lung adenocarcinomas of smokers[28] and is associated with increased metastatic potential in breast cancer[29]. Increased serum lactate dehydrogenase is correlated with tumor stage and tumor burden[30], and cystatin C, a cysteine protease inhibitor expressed in human lung cancers[31], is prognostic in some cancers[32]. The decreased expression of this protease inhibitor may affect the invasive properties of the tumor cell.

The cross-validation analytical strategy we used is particularly informative for these types of gene-expression analyses for disease outcome[33,34], and identification of cross-validated genes with a larger tumor cohort may help refine this risk index for use in a clinical setting. The gene-expression data also provide opportunities to observe overarching patterns that advance our understanding of associations between genes and disease. For example, the top 100 survival genes include those involved in signaling, cell cycle and growth, transcription, translation and metabolism. Expression of many of these genes is likely a function of increased proliferation and metabolism in the more aggressive tumors. Some genes, such as *erb*B2 and *Reg*1A (Fig. 4*a* and *b*), were highly overexpressed in a few patients having poor survival. In one tumor, the *erb*B2 gene was amplified (Fig. 5*a*), demonstrating that genomic changes may underlie the overexpression of a subset of these outlier genes. Immunohistochemistry confirmed protein overexpression in this patient's tumor (Fig. 5*b*). Notably, seven of the eight outlier genes were not amplified, indicating that other mechanisms underlie the increased mRNA expression of these survival-related genes.

Most genes showed a graded relationship between expression and patient survival. Genes such as that encoding VEGF, known to be strongly associated with survival in lung cancer[35,36] were identified as related to patient survival in our study. *VEGF* demonstrated a graded expression pattern, as did the S100P and *crk* oncogene (Fig. 5*b*). S1C0P is a calcium-regulated protein not previously reported in lung cancer. The *crk* gene, the cellular homolog of the v-*crk* oncogene, is a member of a family of adaptor proteins involved in signal transduction and interacts directly with c-jun N-terminal kinase 1 (JNK1)[37]. Although *crk* has not been shown to have a role lung cancer, its role in the MAP-kinase pathway, which leads to activation of matrix metalloproteinase secretion and cell invasion[38], indicates potential involvement in the the tumor cell invasion or metastasis of some lung adenocarcinomas. Among the many genes identified in this study, like *crk*, that may be causally involved in lung cancer progression (Table 1), some were related to survival in many patients, and others in only smaller subsets of patients. This result is consistent with the complex molecular architecture of tumors in general, the heterogeneity of lung adenocarcinomas in particular and the multiple mechanisms underlying tumor-cell survival, invasion and metastasis[39].

Our results demonstrate that a gene-expression risk profile—based on the genes most associated with patient survival—can distinguish stage I lung adenocarcinomas and differentiate prognoses. The particular genes that define the clusters, or are associated with survival, likely reflect the characteristics of the particular tumors included in the analysis. Current therapy for patients with stage I disease usually consists of surgical resection without adjuvant treatment[2,3]. Clearly, the identification of a high-risk group among patients with stage I disease would lead to consideration of additional therapeutic intervention for this group, possibly leading to improved survival of these patients.

Methods

Patient population. Sequential patients seen at the University of Michigan Hospital between May 1994 and July 2000 for stage I or stage III lung adenocarcinoma were evaluated for this study. Consent was received and the project was approved by the local Institutional Review Board. Primary tumors and adjacent non-neoplastic lung tissue were obtained at the time of

surgery. Peripheral portions of resected lung carcinomas were sectioned, evaluated by a study pathologist and compared with routine H&E sections of the same tumors, and utilized for mRNA isolation. Regions chosen for analysis contained a tumor cellularity greater than 70%, no mixed histology, potential metastatic origin, extensive lymphocytic infiltration or fibrosis. Tumors were histopathologically divided into two categories based on their growth pattern: bronchial-derived, if they exhibited invasive features with architectural destruction, and bronchioloalveolar, if they exhibited preservation of the lung architecture. All stage I patients received only surgical resection with intra-thoracic nodal sampling and no other treatments. Stage III patients received surgical resection plus chemotherapy and radiotherapy.

Gene-expression profiling and K-*ras* mutation analysis. RNA isolation, cRNA synthesis and gene-expression profiling were performed as described[24]. Details of gene annotation and K-*ras* mutation analysis are provided in supplementary information.

Northern-blot analysis. Total cellular RNA (10 μg) was separated in 1.2% agarose-formaldehyde gels and vacuum-transferred to Gene Screen Plus (NEN Life Science Products, Boston, Massachusetts). Hybridization conditions and probe labeling were as described[40]. Individual sequence-validated cDNA image clones for human *IGFBP3* (clone 1407750), *LDH-A* (clone 2420241), cystatin C (CTS3; clone 949938) were from Research Genetics (Huntsville, Alabama). The human histone H4 cDNA and the *28S* ribosomal RNA 26-mer oligonucleotide probe were prepared and labeled as described[40].

Gene-amplification analysis. 11 genes were selected for the analysis of genomic alterations. Primers were designed using PrimerSelect 4.05 Windows 32 software (DNASTAR, Madison, Wisconsin), avoiding pseudogenes or potential homologous regions. Forward and reverse primers for the genes are provided (Supplementary Methods online). Quantitative genomic-PCR was then applied and analyzed as described[41].

Immunohistochemical staining. The H&E-stained slides of all primary lung tumors were used to identify the most representative regions of each tumor and a tissue microarray (TMA) block was constructed as described[42]. Immunohistochemistry (IHC) was performed using both routine and sections from the TMA block as described[24]. Detailed methods and the concentrations used for all antibodies are provided in the Supplementary Methods.

Statistical methods. *t*-tests were used to identify differences in mean gene-expression levels between comparison groups. Agglomerative hierarchical clustering[43] was applied using the average linkage method to investigate whether there was evidence for natural groupings of tumor samples based on correlations between gene-expression profiles. To investigate the robustness of the clustering inference, gene-expression values were perturbed by adding random Gaussian error of magnitude obtained from a duplicate sample to each data point and then reclustered to determine concordance in the tumor's class membership. Pearson, χ^2 and Fisher's exact tests were used to assess whether cluster membership was associated with physical and genetic characteristics of the tumors.

To determine whether gene-expression profiles were associated with variability in survival times, 2 separate but complementary approaches were used. In the first approach, the 86 tumors were randomly assigned to equivalent training and testing sets consisting of equal numbers of stage I and III tumors in order to validate a novel risk-index function that captured the effect of many genes at once. In the second approach, cross-validation[44] was used to more robustly identify the genes associated with survival. Briefly, a 'leave-one-out' cross-validation procedure in which 85 of the 86 tumors (the training set) was used to identify genes that were univariately associated with survival. The risk index was defined as a linear combination of the gene-expression values for the top genes identified by univariate Cox proportional-hazard regression modeling[45], weighted by their estimated regression coefficients. Kaplan–Meier survival plots and log-rank tests were then used to assess whether the risk-index assignment to high/low categories was validated in the test set. A more detailed description is provided (Supplementary Methods online).

Note: Supplementary information is available on the Nature Medicine website.

Acknowledgments

We thank D. Sanders for technical assistance; D. Sing for assistance with the figures; and G. Omenn for critical reading of this manuscript. This work was supported by National Cancer Institute grant: U19 CA-85953 and the Tissue Core of the University of Michigan Comprehensive Cancer Center (NIH CA-46952).

Competing interests statement

The authors declare that they have no competing financial interests.

RECEIVED 5 APRIL; ACCEPTED 14 JUNE 2002

1. Fry, W.A., Phillips, J.L. & Menck, H.R. Ten-year survey of lung cancer treatments and survival in hospitals in the United States. *Cancer* 86, 1867–1876 (1999).
2. Williams, D.E. *et al.* Survival of patients surgically treated for stage I lung cancer. *J. Thorac. Cardiovasc. Surg.* 82, 70–76 (1981).
3. Palmero, P.C. *et al.* Postsurgical stage I bronchogenic carcinoma: Morbid implications of recurrent disease. *Ann. Thorac. Surg.* 38, 331–338 (1984).
4. Naruke, T. *et al.* Prognosis and survival in resected carcinoma based on the new international staging system. *J. Thorac. Cardiovasc. Surg.* 96, 440–447 (1988).
5. Kaisermann, M.C. *et al.* Evolving features of lung adenocarcinoma in Rio de Janeiro, Brazil. *Oncol. Rep.* 8, 189–192 (2001).
6. Roggli, V.L. *et al.* Lung cancer heterogeneity: A blinded and randomized study of 100 consecutive cases. *Hum. Pathol.* 16, 569–579 (1985).
7. Gail, M.H. *et al.* Prognostic factors in patients with resected stage I non-small cell lung cancer: A report from the Lung Cancer Study Group. *Cancer* 54, 1802–1813 (1984).
8. Takise, A. *et al.* Histopathologic prognostic factors in adenocarcinomas of the peripheral lung less than 2 cm in diameter. *Cancer* 61, 2083–2088 (1988).
9. Ichinose, Y. *et al.* Is T factor of the TMN staging system a predominant prognostic factor in pathologic stage I non-small cell lung cancer. *J. Thorac. Cardiovasc. Surg.* 106, 90–94 (1993).
10. Harpole, D.H. *et al.* A prognostic model of recurrence and death in stage I non-small cell lung cancer utilizing presentation, histopathology, and oncoprotein expression. *Cancer Res.* 55, 51–56 (1995).
11. Rodenhuis, S. *et al.* Mutational activation of the K-*ras* oncogene: A possible pathogenic factor in adenocarcinoma of the lung. *N. Engl. J. Med.* 317, 929–935 (1987).
12. Slebos, R.J.C. *et al.* K-*ras* oncogene activation as a prognostic marker in adenocarcinoma of the lung. *N. Engl. J. Med.* 323, 561–565 (1990).
13. Horio, Y. *et al.* Prognostic significance of p53 mutations and 3p deletions in primary resected non-small cell lung cancer. *Cancer Res.* 53, 1–4 (1993).
14. Kern, J.A. *et al.* C-*erbB*-2 expression and codon 12 K-*ras* mutations both predict shortened survival for patients with pulmonary adenocarcinomas. *J. Clin. Invest.* 93, 516–520 (1994).
15. Ebina, M. *et al.* Relationship of p53 overexpresson and up-regulation of proliferating cell nuclear antigen with the clinical course of non-small cell lung cancer. *Cancer Res.* 54, 2496–2503 (1994).
16. Mehdi, S.A. *et al.* Prognostic markers in resected stage I and II non-small cell lung cancer: an analysis of 260 patients with 5 year follow-up. *Clin. Lung Cancer* 1, 59–67 (1997).
17. Schneider, P.M. *et al.* Multiple molecular marker testing (p53, c-Ki-*ras*, c-erbB-2) improves estimation of prognosis in potentially curative resected non-small cell lung cancer. *Br. J. Cancer* 83, 473–479 (2000).
18. Herbst, R.S. *et al.* Differential expression of E-cadherin and type IV collagenase genes predicts outcome in patients with stage I non-small cell lung carcinoma. *Clin. Can. Res.* 6, 790–797 (2000).
19. Liotta, L. & Petricion, E. Molecular profiling of human cancer. *Nature Rev. Genet.* 1, 48–56 (2000).
20. Golub, T.R. Editorial: Genome-wide views of cancer. *N. Engl. J. Med.* 344, 601–602 (2001).
21. Bhattacharjee, A. *et al.* Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* 98, 13790–13795 (2001).
22. Garber, M.E. *et al.* Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci. USA* 98, 13784–13789 (2001).
23. Mills, N.E. *et al.* Increased prevalence of K-*ras* oncogene mutations in lung adenocarcinoma. *Cancer Res.* 55, 1444–1447 (1995).
24. Giordano T.J. *et al.* Organ-specific molecular classification of lung, colon and ovarian adenocarcinomas using gene expression profiles. *Am. J. Pathol.* 159, 1231–1238 (2001).
25. Albertson, D.G. *et al.* Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nature Genet.* 25, 144–146 (2000).
26. Kansra, S. *et al.* IGFBP-3 mediates TGF β1 proliferative response in colon cancer cells. *Int. J. Cancer* 87, 373–378 (2000).
27. Vadgama J.V. *et al.* Plasma insulin-like growth factor-I and serum IGF-binding protein 3 can be associated with the progression of breast cancer, and predict the risk of recurrence and the probability of survival in African-American and Hispanic

women. *Oncology* 57, 330–340 (1999).
28. Volm, M., Mattern, J. & Stammler, G. Up-regulation of heat shock protein 70 in adenocarcinoma of the lung in smokers. *Anticancer Res.* 15, 2607–2609 (1995).
29. Clocca, D.R. *et al.* Heat shock protein hsp70 in patients with auxillary lymph node-positive breast cancer:prognostic implications. *J. Natl. Cancer. Inst.* 85, 570–574 (1993).
30. Rotenberg, Z. *et al.* Total lactate dehydrogenase and its isoenzymes in serum of patients with non-small cell lung cancer. *Clin. Chem.* 34, 668–670 (1988).
31. Krepela, E. *et al.* Cysteine proteases and cysteine protease inhibitors in non-small cell lung cancer. *Neoplasma* 45, 318–331 (1998).
32. Kos, J. *et al.* Cysteine proteinases and their inhibitors in extracellular fluids: Markers for diagnosis and prognosis in cancer. *Int. J. Biol. Markers* 15, 84–89 (2000).
33. Golub, T.R. *et al.* Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537 (1999).
34. Hedenfalk, I. *et al.* Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* 344, 539–548 (2001).
35. Ohta, Y. *et al.* Vascular endothelial growth factor and lymph node metastasis in primary lung cancer. *Br. J. Cancer.* 76, 1041–1045 (1997).
36. Shibusa, T., Shijubo, N. & Abe, S. Tumor angiogenesis and vascular endothelial growth factor expression in stage I lung adenocarcinoma. *Clin. Cancer Res.* 4, 1483–1487 (1998).
37. Girardin, S.E. & Yaniv, M. A direct interaction between JNK1 and CrkII is critical for Rac1-induced JNK activation. *EMBO J.* 20, 3437–3446 (2001).
38. Liu, E. *et al.* The Ras-mitogen-activated protein kinase pathway is critical for the activation of matrix metalloproteinase secretion and the invasiveness in v-crk-transformed 3Y1. *Cancer Res.* 60, 2361–64 (2000).
39. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* 100, 57–70 (2000).
40. Hanson, L.A. *et al.* Expression of the glucocorticoid receptor and K-*ras* genes in urethan-induced mouse lung tumors and transformed cell lines. *Exp. Lung. Res.* 17, 371–387 (1991).
41. Lin, L. *et al.* A minimal critical region of the 8p22-23 amplicon in esophageal adenocarcinomas defined using STS-amplification mapping and quantitative PCR includes the GATA-4 gene. *Cancer Res.* 60, 1341–1347 (2000).
42. Kononen, J. *et al.* Tissue microarrays for high throughput molecular profiling of tumor specimens. *Nature Med.* 4, 844–847 (1998).
43. Johnson, R. & Wichern, D.W. *Applied Multivariate Statistical Analysis.* 543–578 (Prentice Hall, New Jersey, 1988).
44. Stone, M. Asymptomics for and against cross-validation. *Biometrika* 64, 29–38 (1977).
45. Cox, D.R. Regression models and life tables. *J.R. Stat. Soc.* 34, 187–220 (1972).